Title	ランダム行列理論との比較によるNYSE株価1時間変動の解析(2)(経済物理学とその周辺,統計数理研究所研究会共同研究集会,経済物理学2009-ミクロとマクロの架け橋-,京都大学基礎物理学研究所2009年度前期研究会,研究会報告)
Author(s)	田中, 美栄子; 伊藤, 大哲; 田中, 瑶子; 木戸, 丈剛
Citation	物性研究 (2010), 93(5): 677-678
Issue Date	2010-02-05
URL	http://hdl.handle.net/2433/169223
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

ランダム行列理論との比較による NYSE 株価 1 時間変動の解析(2)

田中 美栄子[†], 伊藤大哲, 田中 瑶子, 木戸丈剛
鳥取大学大学院工学研究科[†], 鳥取大学工学部知能情報工学科

1. はじめに

時系列間の同時刻相関行列の固有値分布を調べ,ランダム行列から作った相関行列の固有値分布からの外れ成分を有意成分と見なすことより,ランダム性の強い時系列から有意成分を取り出す手法[1]を 2002 年の NYSE 株価の中で,取引量の多い上位 569 社の株価 tick データに適用した結果を,文献[2]の 1994 年の同データに対する結果と比較しつつ報告する.

2. 株価 tick データの同時刻相関行列

等間隔でない tick データから同時刻相関行列を作るため,定時に最も近い時点での価格を代表値として用いる.10 時から 15 時まで,1 日あたり6点に 2002 年の営業日 252 日をかけた長さ T=1512 の時系列を N=569 社に対して抽出した時系列の行列 $x_{i,k}$ (1,..N; k=1,..,T) から作った相関行列

$$C_{i,j} = \frac{1}{T}\sum_{k=1}^{T} x_{i,k} x_{j,k} \qquad (1)$$

の固有値スペクトルを,ランダム行列理論のよく知られた結果と比較して有意成分を分離する.

2-1. 変数の確認

株価 $S_{i,k}$ から作った対数収益率

$$X_{i,k} = \ln(S_{i,k+1}/S_{i,k}) \approx \Delta S_{i,k}/S_{i,k} \qquad (2)$$

から平均値を差し引き標準偏差で割って正規化した変数

$$x_{i,k} = \frac{X_{i,k} - \langle X_i \rangle}{\sigma} \qquad (3)$$

を(1)式の変数として用いる.1994 年のデータについてはこの変数の頻度分布は正規分布と非常に近く,ランダム行列理論と比較するのに良い変数であると言えたが,2002 年のデータに対しては大きく外れる.

Fig.2 に 1994 年と 2002 年の NYSE-TAQ (trade)の株価時系列から作成した正規化対数収益率（角）とその最適関数淡）と標準正規分布（濃）の比較を示し,(b)に対応するの結果を並べて示す.1994 年では(4)式の変数の頻度分布が正規分布にほぼ一致するのに対し,2002 年では大きく異なる.これにより,2002 年には(4)式の変数が必ずしも良い変数とは言えない可能性が示唆される.

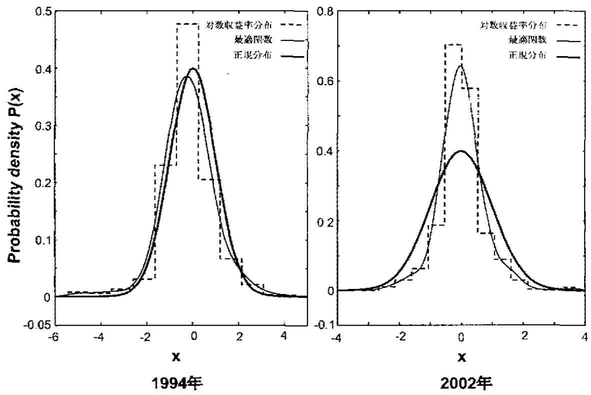

Fig.1 1994 年の収益率は正規分布に近いが、2002 年は異なる

2-2. 相関行列の固有値スペクトル

上記の手続きで(2)式により計算した相関行列 C の固有値 λ と固有ベクトル U は

$$CU = \lambda U \qquad (4)$$

の関係を満たす.C は N×N の成分を持ち,N=569 と大きいので Jacobi 回転を使った数値計算で対角化した.

ランダム時系列の相関行列の固有値分布は

$$N \to \infty,\ T \to \infty,\ Q = T/N = \text{const.} \qquad (5)$$

の極限で理論的に次式が導かれる.

$$P_{rm}(\lambda) = \frac{Q}{2\pi}\frac{\sqrt{(\lambda_{+} - \lambda)(\lambda - \lambda_{-})}}{\lambda} \qquad (6)$$

3. NYSE の tick 株価同時刻相関行列

固有値分布はFig.2 に角グラフで示すように,ランダム行列に重なる部分と外れた部分に分かれる.実曲線で(7)式を示し,それにまとわりつく淡色の角グラフは正規乱数で計算した結果である.固有値が 4 以下の部分はほぼランダム成分と見なせる.4 以上の 10 固有値:λ_1 = 166.4, λ_2 = 20.6, λ_3 = 11.3, λ_4 = 8.6, λ_5 = 7.7, λ_6 = 6.5, λ_7 = 5.8, λ_8 = 5.3, λ_9 = 4.1, λ_{10} = 4.0 が有意成分となる.

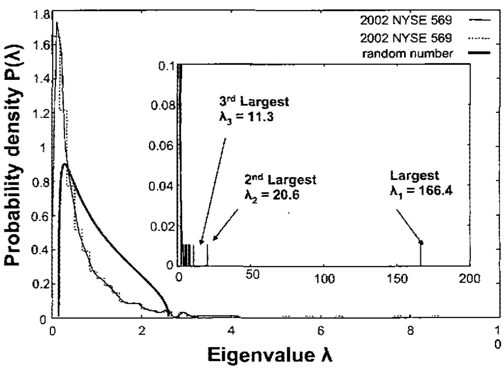

Fig.2 固有値分布(2002)と理論式との比較

次にこの固有値に対する固有ベクトルの成分を調べる.文献[2]では, U_1は業種によらず大企業が並んだが.,今回の 2002 年の結果では金融業と銀行が並んだ.この意味としては, 2002 年に金融業と銀行が巨大化して株式市場の中心となった事が反映している結果と解釈できる.また,1994 年には優勢だった半導体に代わって食品や電気・エネルギー関連株が上位に出ているのが特徴的であるが,これは産業構造が 1994 年から 2002 年の間に変化して,半導体が下火になり,代わって金融,食品,電気・エネルギー株などがNYSEの主力となったことを示唆すると考えられる.

表 1.固有ベクトル(U_k)成分に見る業種の偏り

U_k	業種成分の偏り
U_1	10 社中 6 社が金融関連,2 社が銀行
U_2	10 社中 6 社が食品関連
U_3	10 社全てが電気・エネルギー関連
U_4	10 社中食品 5 社,電気・エネルギー3 社
U_5	10 社全てが電気・エネルギー関連
U_6	10 社中 3 社が電気関連
U_7	10 社中 3 社が食品関連
U_8	10 社中 8 社が小売業
U_9	10 社中 5 社が鉱業関連,3 社は通信
U_{10}	10 社中 8 社が通信関連

4. 株種 N と時系列長 T の検討

第2節に述べたように,1時間毎の代表値を採用すると T=1512 の時系列が作れる.もっと細かく,30 分毎の代表値を採ればTは2倍以上に増えるが取引量の少ない株価は使えなくなるので,Nが減ってしまう.どのくらいのNが適当なのかを見るため,ランダム行列理論から導かれる(7)式が,正規乱数により発生させた時系列の相関行列の固有値分布により支持されるかどうかを N=100,200,400 について調べた.Fig.4 に示すように,N=400 で理論と実験の一致が良くなる.これによりNは 400 以上が望ましいと結論した.

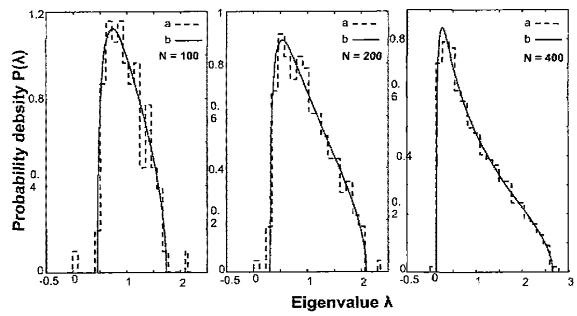

Fig. 3 N = 400 で理論式 Eq.(7)とよく一致する

6. 参考文献

[1]V. Plerou,et.al.,PRE65,066126,2002

[2]田中美栄子,他“ランダム行列との比較による NYSE 株価 1 時間変動の相関行列解析”経済物理 2009 (2009 年 9 月 8 日, 基礎物理学研究所)